韓國地形空間情報學會論文集 第3卷 第1號 1995年 6月 pp. 115-121

격자형 지질정보의 자료유도 통합을 위한 이론적 배경

Theoretical Background for Data-driven Integration of Raster-based Geological Information

李 箕 遠* Lee, Kiwon

智 光 薰** Chi, Kwang-Hoon

要 旨

최근 지리정보시스템의 여러 지질학적 응용중에서 광물탐사를 위한 격자형자료의 공간적 통합론에 관한 연구가 많이 이루어지고 있다. 본 연구에서는 보통 확률, 통계적 배경을 갖는 목표유도형방법과 구분되는 자료유도형 방법의 예로서 Dempster-Shafer의 이론과 퍼지이론의 이론적 배경을 자료재표현의 원리와 자료통합논리에 입각하여 설명하고자 한다. 기존의 지질, 지화학및 물리탐사정보를 이용한 사례 연구에서 위의 두 이론은 광물탐사문제에 상당히 유용한 결정보조 정보를 제공하는 것으로 입증되고 있으며, 본 연구에서 논의된 몇 가지 관련 사항들은 이 이론들의 보다 적절한 실제 적용및 해석에 도움이 될 것으로 생각된다.

ABSTRACT

Recently, spatial integration for mineral exploration is regarded as an important task of various geological applications of GIS. Therefore, theoretical bases of data representation and reasoning concerned with Dempster-Shafer theory and Fuzzy theory were systematically explained as the data-driven integration methodologies for raster-based geoinformation; they are distinguished from target-driven methodology based on statistical background. According to previous actual applications of these methods to mineral exploration, they have been proven to provide useful information related to hidden target mineral deposits, and it is thought that some suggestions in this study are helpful to further real applications including representation, reasoning, and interpretation stages in order to obtain a decision-supporting layer.

1. 서 론

현재 다양한 지질정보를 지리정보시스템(Geographical Information Systems, GISs)을 이용하여 응용하기 위한 여러 가지 시도가 연구되고 있다. 본 연구에서는 GIS의 여러 가지 지질학적 응용 방안중에서도 중요한 과제로 인식되고 있는 공간자료 통합에 관한 이론적 배경을 설명하고자 한다.

현재까지 개발된 대부분의 상업용 또는 비상업용 지리정보시스템에서 제공되는 대부분의 자료통합모듈은 부울집합론(Boolean Set Theory)에 근거한 이진수 화상처리기법 (Image Processing technique)에 근간을

* 서울대학교 지구과학교육과 박사과정

** 한국자원연구소 환경지질 선임연구원

두고 있으므로, 지화학과 물리탐사자료가 포함된 공간적 지질정보에 대하여 곧바로 이러한 방법을 적용하는 경우에는 통합 결과로 제시되는 결정 보조 정보(decision - supporting layer)의 배경이 충분치 못하거나, 또한 경우에 따라서는 심각한 해석상의 오류를 야기하여 예측될 수 있는 결과와 큰 차이를 보일 수도 있다. 따라서 다양한 공간적 지질정보를 다루기 위한 수학논리적 자료 통합론의 필요성이 부각되고 있다. 이러한 배경을 가지고 그간 개발된 통합이론은 기존에 밝혀진 확실한 탐사대상에 관한 정보로 부터 유도되는 방법(Target-driven methodology)과 탐사 목표에 대한 GIS의 공간상의 정보가 거의 없거나, 불확실한 경우에 시도될 수있는 자료유도방법(Data-driven methodology)의 두가지로 크게 구분할 수있다. 예를 들어 Bayes' theorem을 이용한 확률적 방법, weight of evdience 방법(Bonham - Carter et al.,1988), decision-tree 방법(Reddy and Bonham-Carter, 1990) 등은 전자의 경우에 해당하고, Dempster-shafer 의 이론을 이용한 방법(Lee and Richards, 1987; Moon, 1990)과 퍼지이론을 이용한 방법(An et al., 1991 ; Eddy et al., 1994)은 후자에 해당한다. 한편 실제 적용에서는 위의 대부분 방법들을 이용한 자료통합의 목적이 대부분 유용 광물 탐사분야에 집중되어 있으므로, 본 연구에서 논의되는 이론적 배경도 광물탐사를 위한 실제 자료통합문제에 적용시킬수 있도록 하였다.

따라서 본 연구의 의의는 차후의 국내의 지화학, 물리탐사자료를 이용한 자료통합 과제를 수행하기 위한 이론적 배경을 제공하는 데 있으며, 공간적 사고(Spatial reasoning)를 통하여 자료유도방법으로 적용될 수있는 Dempster-Shafer(D-S) 이론과 퍼지이론을 지질정보통합론의 관점에서 Moon(1993)의 기존의 연구를 바탕으로 하여 종합적으로 설명하였고, 이러한 방법들의 실제 적용에서 발생하는 문제에 대한 제안점을 논의하였다.

2. 자료통합론

격자 형태의 총 n개의 지질학적 자료가 있을 때, GIS의 기본도(Base map)로 표현되는 탐사 지역내에서 각각의 자료는 $E_k(k=1,\cdots,n)$로 나타낼 수 있다. 여기서 E_k로 표현된 자료들은 Resampling, Interpolation, Georefemcing등과 같은 기본적인 GIS 단계를 거친 지화학, 물리탐사자료를 의미한다.

탐사 목표(Exploration target)를 일단 ET로 정의하면, 탐사목적은 실제 탐사 지역내에서 부존하는 특정한 유용광물광상의 분포를 찾기 위한 것이고, 여기서 광물자원의 분포는 기존에 발견된 광상및 미발견된 광상의 기본도내의 실제 위치를 의미한다. 본 연구에서 언급하는 자료유도방법론에서는 결정보조정보를 얻기 위하여 $E_k(k=1,\cdots,n)$만을 고려한다. 한편 E_k내의 자료를 적용되는 자료유도 통합이론에 맞도록 재표현(Representation)하기 위하여, E_k를 구성하는 각각의 속성값 또는 실제 관측값 e는 다음과 같은 수학적 mapping을 통하여 재표현한다.

$$d_k(e): E_k \rightarrow [\ 0,1]$$

부울집합론에서는 이 mapping을 통하여 이진수의 의미로 존재의 유무만을 지시하나, 지질정보를 다루는 경우에는 탐사목적을 위하여 지질정보를 mapping하는 과정에서 0과 1사이의 값들이 실제로 중요한 의미를 갖게 된다. 따라서 지질정보통합론은 이러한 mapping을 통한 자료의 재표현과 통합(Integration)의 법칙을 반드시 함께 다루어야 하는데, Dempster-Shafer의 이론과 퍼지이론은 이러한 두 가지 사항을 효과적으로 처리할 수 있는 중요한 방법으로 간주될 수 있다.

2.1 Dempster-Shafer의 이론를 이용한 자료통합론

이 방법은 자료의 표현방법으로 bpa(Basic probability assignment)함수를 통한 확률적인 표현 방법을 이용하지만, 확률통계적인 배경과는 기본적인 개념이 다르다. D-S이론에서는 실공간 S로 부터 치환공

간 θ로의 다중치 mapping을 표현하기 위하여 아래의 관계식을 이용한다.

$$\Gamma : S \rightarrow 2^{\theta}$$

또한 위의 기본 관계에서 S내의 한 요소 s가 존재할 때, 이에 호환적으로 대응하는 θ내의 한 요소 t가 동시에 존재할 수 있다면, S의 상, G(s)는 다음과 같이 표현 될 수있다.

$$G(s) = \{ t | t \in \theta \}$$

이러한 기본관계로 부터 D-S이론에 입각하여 자료의 재표현을 하는데 직접 관계되는 bpa함수는 아래의 관계식으로 표현된다.

$$m(A) = \frac{\sum_{G(s_i)=A} p(s_i)}{1 - \sum_{G(s_i)=\emptyset} p(s_i)}$$

이때 A는 관심요소 (focal element)를 의미하며, 지질도내의 특정한 암종이나 암상 또는 적절한 군집화과정(Clustering)을 통한 지화학자료나 물리탐사자료내의 임의의 군집을 의미한다. 그러나 실제 치환역 θ의 확률분포를 결정하는 요소로 위의 관계식으로 부터 도출되는 bpa함수외에 실제 mapping의 목표가 되는 명제를 표현하는 부가적인 집합 B가 필요하며, 이때 B는 실제의 탐사 목표가 되는 특정유용광상을 의미한다.

$$Spt(B) = Spt(ET|e) = \sum_{A \subset B} m(A)$$
$$Pls(B) = Pls(ET|e) = \sum_{A \cap B \neq \emptyset} m(A)$$

이 이론에 의하여 D-S이론의 Support function을 의미하는 $Spt_k(ET|e)$는 탐사 목표에 부합되도록 할당되는 모든 bpa함수의 합이 되는 조건부 확률의 하한 경계로서, Plausibility function를 나타내는 $Pls_k(ET|e)$는 조건부 확률의 상한 경계로 해석될 수 있다. 한편 아래의 관계식들을 통하여 상호 관련성을 살펴 볼 수 있다.

$$Spt_k(ET|e) + Spt_k(\sim ET|e) \leq 1$$
$$Pls_k(ET|e) = 1 - Spt_k(\sim ET|e)$$
$$Pls_k(ET|e) \geq P_k(ET|e) \geq Spt_k(ET|e)$$

한편 이러한 관계는 일종의 불확실성을 의미하는 신뢰역으로 $[\, Spt_k(ET|e), Pls_k(ET|e)]$ 의 간격을 갖게되는데, 이 간격의 차는 D-S이론에 따른 또 다른 함수, Ignorance function, Ign_k으로 다음과 같이 표현될 수있다.

$$Pls_k - Spt_k = Ign_k$$

따라서 만약 탐사 목표에 대한 수학적인 의미로 절대적인 가치를 갖는 정보가 되는 관측치가 존재한다면 다음과 같은 관계식이 성립한다.

$$Pls_k = P_k = Spt_k$$

Lee and Richards(1987)는 D-S이론의 실제 응용으로 원격탐사자료의 목표지향적 격자분류화(classification)에 이 이론을 적용할 수있는 가능성을 제시하면서, Spt, Pls, 및 Ign을 실제 자료상의 각각의 격자 간격에 할당하는 사례연구를 수행한 바있고, Moon (1990)은 지질정보와 물리탐사자료를 통합하기 위하여 D-S이론의 실제 적용에 전문가의 주관적경험에 의하여 표현된 세 함수의 mapping을 통하여 광물탐사를 목적으로 한 지질, 물리탐사자료의 통합을 광물탐사목적으로 실시한 바 있다. 한편 D-S의 이론에 입각한 자료의 통합은 두 bpa함수로 부터 새로운 bpa함수를 구하는 과정을 의미한다.

$$m_1 \oplus m_2(C) = \frac{\sum_{A_i \cap B_j = C} m_1(A_i) m_2(B_j)}{1 - \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j)}$$

실제 자료통합문제에서는 위의 기본관계로 부터 유도되는 다음의 관계식이 직접 이용될 수 있다. 만약 주어진 탐사 목표에 대한 다양한 자료들이 다음과 같이 취합되어 있다면,

$$\{Spt_1(ET|e_1), Spt_2(ET|e_2), \cdots, Spt_n(ET|e_n)\}$$
$$\{Pls_1(ET|e_1), Pls_2(ET|e_2), \cdots, Pls_n(ET|e_n)\}$$

$Spt_i(ET|e_i)$과 $Spt_j(ET|e_j)$, $Pls_i(ET|e_i)$과 $Pls_j(ET|e_j)$에 대하여 다음과 같은 combination rule 이 성립한다.

$$Spt(ET|e_i, e_j) = \frac{ab + a(1-b-b') + b(1-a-a')}{1-ab'-a'b}$$

$$Pls(ET|e_i, e_j) = 1 - \frac{a'b' + a'(1-b-b') + b'(1-a-a')}{1-ab'-a'b}$$

$$a = Spt_i(ET|e_i)$$
$$a' = 1 - Pls_i(ET|e_i)$$
$$b = Spt_j(ET|e_j)$$
$$b' = 1 - Pls_j(ET|e_j)$$

따라서 n개의 모든 자료를 처리하기 위해서 위의 관계식을 n-1번 수행된 뒤 최종 결과를 얻을 수 있다. Chung and Moon (1990)은 광물탐사를 목적으로 한 모형연구를 통하여 지질정보통합론에서 확률적 통합방법과 D-S이론을 직접 비교하여 D-S이론의 타당성 및 효용성을 입증한 바 있다. 또한 Leclerc and Chung (1993)은 Dempster-Shafer의 이론을 PC-based GIS에 적용할 수 있는 FAVMOD를 개발한 바 있으며, 이를 실제 자료에 적용시키는 데 있어서는 본 연구에서 개관된 이론적 배경이 필요하다.

2.2 퍼지 이론을 이용한 자료 통합론

퍼지이론은 퍼지집합론(Fuzzy set theory)과 퍼지논리(Fuzzy logic)의 두 측면을 기본하여 성립된다. 퍼지집합론은 mapping을 위한 membership 함수의 형태로 원래의 자료를 재표현하게 된다. 퍼지논리는 실제 자료의 통합문제를 해결하기 위한 퍼지연산을 의미한다.

따라서, $U_k(ET|e)$을 membership 함수라 할때 다음과 같은 관계가 성립한다.

$$U_k(ET|e_j) = d_k(e_j), \quad j = 1, \cdots, m$$

윗 식에서 j는 E_k내의 탐사목표에 따라 분류된 군집을 의미하며, m은 E_k내의 탐사목표에 대한 총 군집수를 뜻한다.

이러한 membership function은 여러 가지의 기존에 알려진 연속, 불연속 함수을 이용하여 표현될 수 있다. 자연과학에서 이용되는 membership함수의 형태는 Delta함수, Step함수, Ramp함수, S함수등과 같이 단순한 함수가 이용될 수 있으나, 아래와 같은 대칭 혹은 비대칭형의 Bell형 함수가 보편적으로 이용되고 있다.

대칭형 Bell형 함수의 경우,

$$U_k(ET|e_j) = \frac{1}{\{f + g(e_j - h)^2\}}$$

여기서 f는 탐사목표에 대한 최대 가능률(possibility)로 보통의 경우 1을 이용한다. 한편 g와 h는 모형변수를 의미한다.

비대칭형 Bell형 함수의 경우,

$$e_j < h \text{ 의 경우}, \quad U_k(ET|e_j) = \frac{1}{\{f + g(e_j - h)^2\}}$$

$$e_j \geq h \text{ 의 경우}, \quad U_k(ET|e_j) = f$$

위의 Bell형 함수들은 일차원적 문제에서는 쉽게 적용될 수 있으나, 이차원적, 공간상의 지질자료, 특히 지화학, 물리탐사자료의 경우 Histogram분석을 통하여 이러한 정보의 공간적 특성을 파악한 뒤, 적절한 탐사목표에 따른 Clustering을 통하여 membership 함수를 구할 수도 있다. 그러나 퍼지이론을 실제의 자연과학적 또는 공학적 문제에 적용하는 데 있어서 최적의 membership function을 결정하는 문제는 쉬운 문

제가 아니지만 이 퍼지이론 자체는 실제 문제 해결에 있어서 기존의 확률 및 통계적 배경에 크게 구애받지 않는다는 중요한 장점은 주지할 사실이다. 따라서 일반적으로 자료통합론에서 membership function의 결정에 중요한 사항은 자료자체를 탐사목표에 대하여 어떻게 현실적으로 관련시킬 수 있느냐 하는 점이라고 할 수 있다. 일단 탐사 목표에 대한 membership function이 결정되고, membership function의 형태가 convex한 함수로 표현 될 수 있다면 다음과 같이 가중치적용방법(Burrough, 1989)이 효과적으로 이용될 수 있다.

$$U_k(ET|e) = \sum_{i=1}^{n} w_i U_k(ET|e_i)$$
$$\sum_{i=1}^{n} w_i = 1,\ w_i > 0$$

이러한 membership 함수를 통합하기 위한 여러 퍼지 연산자중에서 지질정보의 통합에 실제 적용가능한 연산자로는 min연산자, max연산자가 대표적인 것으로 알려져 있으며, 이 두 연산자는 각각 논리적으로 AND와 OR의 의미를 갖는다. 실제 계산과정은 다음과 같이 간단한 연산을 통하여 임의의 p점에 대한 통합결과를 얻게 된다.

$$U(ET|e_1, e_2, \cdots, e_n)_{at\ p} = \min\ \{U_1(ET|e_1), U_2(ET|e_2), \cdots, Un(ET|e_n)\}$$

$$U(ET|e_1, e_2, \cdots, e_n)_{at\ p} = \max\ \{U_1(ET|e_1), U_2(ET|e_2), \cdots, Un(ET|e_n)\}$$

한편 실제 문제에 있어서 위와 유사한 논리적 배경을 갖는 LMin, LMax연산자가 이용될 수도 있다.

$$U(ET|e_i, e_j)_{at\ p} = \min\ (U_1(ET|e_i) + U_2(ET|e_j) - 1, 0\ )$$
$$U(ET|e_i, e_j)_{at\ p} = \max\ (U_1(ET|e_i) + U_2(ET|e_j), 1\ )$$

또다른 유용한 연산자로 다음과 같은 소위 γ 연산자가 이용되기도 한다(An et al., 1991). 이연산자에서 보통 0.975의 γ값이 원래의 이 연산자의 특성에 맞는 안정적인 해를 제공하나, $\gamma = 0$ 이거나 $\gamma = 1$ 인 경우에는 각각 또다른 연산자인 퍼지 산술곱 연산자와 퍼지 산술합 연산자로 각각 호환된다.

$$U(ET|e_1, e_2, \cdots, e_n)_{at\ p} = [\ \prod_{k=1}^{n} U_k(ET|e_k)]^{1-\gamma} - [\ 1 - \prod_{k=1}^{n}(1 - U_k(ET|e_k)]^{\gamma}$$

비록 본 연구의 주요 관심으로 지질정보의 통합문제에서 GIS에 대한 퍼지이론의 응용문제를 설명하였지만, 한편으로 퍼지이론은 본질적으로 정보의 불확실성에 관계된 문제를 효과적으로 다루기 위하여 시작되었기 때문에 GIS의 여러가지 측면에서 응용될 수 있으며(Lam, 1992), 앞으로 많은 연구가 필요할 것으로 생각된다.

3. 토의 및 결론

앞에서 설명한 자료유도형 정보통합방법들은 기본적으로 공간상에서 나타나는 자료의 탐사목표에 대한 속성과 분포 특성들이 통합 결과에 중요하게 작용한다. 실제 지화학자료와 달리 물리탐사 자료는 지표 근처또는 지하 심부에서 발생되는 지질학적 특성과 물성이 지표면상의 이상치의 분포를 결정하는 주요 변수이기 때문에 탐사 목표에 따른 탐사 자료의 특성을 분석하기 위한 연구가 GIS로의 registration에 앞서 선행되어야 한다. 한편 Yen(1983)은 Dempster-Shafer의 이론에 기초하여 퍼지이론을 적용시키는 방안을 제시하였으나, 지질 정보를 통합하는 문제에서는 이 두 방법의 결과를 직접비교하는 연구보다 탐사조건, 자료의 종류및 탐사목적에 따라 적절한 방법을 선택할 수 있는 준거를 마련하는 연구가 더욱 필요한 것으로 생각된다.

본 연구에서 설명한 두 가지의 자료유도통합이론에 대하여 실제 적용문제와 관련된 몇 가지의 사항을 다음과 같이 정리하였다.

1. 자료유도방법에 기초한 자료통합과정은 공간상의

한 격자크기에 대하여 크게 영향을 받지 않으므로, 실제 통합과정에 격자의 크기는 탐사지역내의 분포하는 기존에 발견된 광상의 규모및 산상에 따라 결정되어야 한다. 물론 격자간격을 조밀하게 하거나, quadtree형태로 자료를 구성함으로서 통합된 자료의 분해능을 증가시킬수 있다.

2. 위의 두가지 방법은 기본도내에서 탐사자료의 범위가 일치하지 않는 경우, 또는 자료가 누락되는 경우에도 자료통합이 가능하다. 전통적인 확률, 통계적 접근방법에서는 이용되는 기본자료의 범위가 일치해야 자료통합이 가능하며, 결과의 신뢰도가 증가되는데, D-S이론과 퍼지이론에서는 다소 주관적이기는 하나, 한 자료내에서 정보가 없는 지역에서도 탐사목표에 대한 유의미한 mapping을 허락하므로, 탐사목표에 대한 전체 탐사지역의 통합정보를 얻을 수 있다.

3. D-S이론에서는 확률적 접근에서 가능한 결과의 불확실성처리및 해석보다 구체적인 공간상의 불확실성을 처리할 수있으므로 통합결과의 해석에 중요한 근거를 제시할 수있다. 그러나 이 이론에 의한 결과는 하나의 결정보조론적 통합정보를 얻기 위하여 $Spt, Pls,$ 와 Ign의 측면을 모두 고려해야 하므로 때때로 해석상의 어려움이 따르기도 한다.

4. 사전에 탐사목표에 대한 공간상의 사전정보(a prior information)를 절대적으로 필요로 하지 않는다. 목표유도형방법에서는 이러한 정보가 필수적이기 때문에 부정확한 정보를 가지고 통합하는 경우 심각한 오차를 나타낼 수 있다.

5. 퍼지이론을 이용한 통합방법에서는 어떠한 퍼지 연산자를 적용하는가에 따라 통합정보가 달라지므로 탐사목적에 따라서 연산자의 적용성이 달라진다. 그러므로 적절한 몇 가지의 연산자를 적용한 결과도 역시 각각 의미가 다르므로 몇 가지의 해석이 가능하다.

6. 현재까지 이 두 방법에 대하여 오차전파및 처리에 관한 체계적인 이론이 정립되어 있지 않다. 현재 여러 GIS에서 오차를 효과적으로 분석할 수 있는 모듈이 일부 제공되기는 하나, 대부분의 경우 이러한 모듈은 확률, 통계적인 접근방법을 기본으로 하기 때문에 D-S이론이나 퍼지이론과 같이 possibility를 다루는 문제에는 직접적인 적용이 곤란하다.

참 고 문 헌

1. An, P., Moon. W. M., and Rencz, A. (1991) Application of Fuzzy Set Theory for Integration of Geological, Geophysical and Remote Sensing Data, Canadian Jour. of Exploration Geophysicists, 27, p. 1-11.
2. Bonham-Carter, G. F., Agterberg, F. P. and Wright, D. F. (1988) Integration of Geological Datasetsfor Gold Exploration in Nova Scotia, Photogrammetric Engineering and Remote Sensing, 54(11), p. 1585-1592.
3. Burrough, P.A. (1989) Fuzzy mathematical methods for soil survey and land evaluation, Jour. of Soil Science, 40, p. 477-492.
4. Chung, C. F. and Moon, W. M. (1991) Combination Rules of Geoscience Data for Mineral Exploration, Geoinformatics, 2(2), p. 159-169.
5. Eddy, B. G., Jefferson, C. W., Bonham-Carter, G. F. (1994) Mineral Potential Mapping using a Knowledge-based Fuzzy Logic Approach for Melville Island, Northwest Territories, Geological Survey of Canada, Open File Report 2795, 50p.
6. Lam, S. S. M. (1992) Uncertainty Management with Fuzzy Sets in a Commercial Vector GIS, UCGE Report 20059, 99p.

7. Leclerc, Y. and Chung, C. F. (1992) FAVMOD-Integration of Spatial Geoscience Information, Geological Survey of Canada, Open File Report 2577, 88p.
8. Lee, T. and Richards, J. A. (1987) Probabilistic and Evidential Approaches for Multisource Data Analysis, IEEE Trans, on Geoscience and Remote Sensing, 25(3), p. 283-293.
9. Moon, W. M. (1990) Integration og Geophysical and Geological Data using Evidential Belief Function, IEEE Trans. on Geoscience and Remote Sensing, 28(4), p. 711-720.
10. Moon, W. M. (1993) On Mathematical Representation and Integration of Multiple Spatial Geoscience Data Sets, Canadian Jour. of Remote Sensing, 19(1), p. 63-67.
11. Reddy, R. K. T. and Bonhan-Carter, G. F. (1991) A Decision-tree Approach to Mineral Potential Mapping in Snow Lake Area, Manitoba, Canadian Jour. of Remote Sensing, 17(2), p. 191-200.
12. Yen, J. (1983) Generalizing the Dempster - Shafer Theory to Fuzzy Sets, IEEE Trans. on Systems, Man, and Cybernetics, 20, p. 559-585.